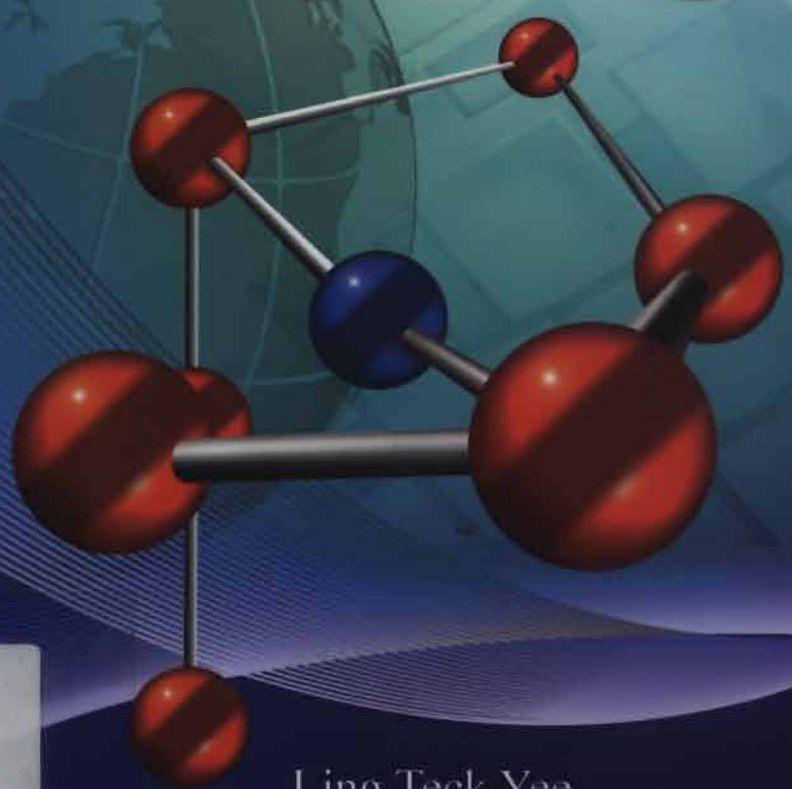
Elementary Statistics *for* Sciences *with* SPSS



Ling Teck Yee
Hong Kian Sam

UNIVERSITI MALAYSIA SARAWAK
UNIMAS

Elementary Statistics for Sciences with SPSS

Ling Teck Yee
Hong Kian Sam

Universiti Malaysia Sarawak
2012

First Publication 2012

Published in Malaysia by
Universiti Malaysia Sarawak,
94300 Kota Samarahan,
Sarawak, Malaysia.

Printed in Malaysia by,
Lee Ming Press Sdn Bhd (541980-U)
No. 48, Ellis Road,
93300 Kuching, Sarawak
Malaysia

Perpustakaan Negara Malaysia Cataloguing-in-Publication Data

Ling, Teck Yee
Elementary statistic for sciences with SPSS / Ling Teck Yee & Hong Kian Sam.
Bibliography: p. 118
ISBN 978-967-5527-37-1
1. SPSS for Windows. 2. Social sciences-Statistical methods-Computer programs. 3. Mathematical statistics. I. Hong, Kian Sam. II. Title.
519.502855361

PREFACE

Many undergraduate students, postgraduate students and academic staff are engaged in quantitative based research in their research projects, research theses or research grants. In addition, many of them use statistical softwares such as SPSS (Statistical Packages for the Social Sciences) and SAS (Statistical Analysis System) as they analyse their quantitative data.

With the widespread availability of statistical softwares and the extensive application of quantitative methods in many academic fields, statistical courses now emphasize statistical reasoning more than computational skills. Questions of how have given way to the more challenging questions of why, when and what.

This book arises from the workshop materials used in the workshop on "Elementary Statistics using SPSS" conducted by the Research and Innovation Management Centre, Universiti Malaysia Sarawak. The goal of this book is to provide a set of self-paced exercises with examples from the sciences that readers can work independently, learning the software skills of SPSS while understanding the underlying statistical concepts and techniques.

This book covers statistical concepts commonly found in elementary statistics courses covering both descriptive statistics and inferential statistics. The goals of this book are to assist the readers understand the concepts and techniques of statistical analyses and use the SPSS for data analyses. It is not meant to replace statistics textbooks on elementary statistics.

The authors would like to acknowledge the assistance of the Research and Innovation Management Centre and the Publication Division for their support in preparing and publishing the book. We would also like to thank Mr George Tan Geok Shim and Miss Gan Siew Ling for testing the SPSS sessions in the text and all the workshop participants who have contributed towards the improvement of the content and presentation of the book.

Associate Professor Dr Ling Teck Yee
Associate Professor Dr Hong Kian Sam

FOREWORD

At the heart of a university's mission is a commitment to intellectual discovery and development. Not only will research-led teaching benefit students, through research, development, innovation, and dissemination, society gains cultural, social, economic and political benefits. Furthermore, a university's reputation is directly linked to the quality of its research. As its reputation in research grows, so does its ability to attract and retain high-performing staff and outstanding research students. Research excellence is the key to achieving a university's aspiration for leadership as a premier university and gaining recognition as a world-class institution.

Research is conducted in an increasingly globalised and interdisciplinary context, using sophisticated techniques and technological tools. Generally, early career academics require guidance and assistance in their research endeavors to contribute to the research, development, innovation and publication activities of a university.

Universiti Malaysia Sarawak, through the Research and Innovation Management Centre has implemented various measures to enhance the capacity of early career academics and lecturers to carry out these core activities, among which is the Research Capacity Enhancement Workshop Series. The workshop "Elementary Statistics using SPSS for Sciences" is an example of one such workshop which has been well-received by the participants.

In another step forward, the Research and Innovation Management Centre in collaboration with the workshop facilitator has edited the workshop materials into a book. This book will provide another resource in improving the quality of quantitative research carried out at the university and it is appropriate for early career researchers and undergraduate and postgraduate students.

I would like to take this opportunity to congratulate the authors and commend the Research and Innovation Management Centre for their hard work and I look forward to more such efforts in the future.

Professor Dr Peter Songan
Deputy Vice-Chancellor (Research And Innovation)

FOREWORD

Universiti Malaysia Sarawak believes research complements and invigorates teaching, as well as collectively supports the preservation, dissemination and advancement of knowledge for the betterment of society as the central tenet of its mission. The Research and Innovation Management Centre was established in 2005. The Research and Innovation Management Centre is responsible for overseeing the management of research in the university. Among its many responsibilities are the tasks of assisting lecturers to secure research grants, promote research culture in the university, and organise relevant training programs to enhance the university's staff capability in research and development activities.

Since 2009, the Research and Innovation Management Centre has implemented a series of Research Capacity Enhancement workshops to promote research and publication activities among early career academics at the university. In this workshop series, the participants learn to write writing successful research grant proposals, apply various research methodologies to their research, and carry out data analyses with the aid of computers and using the computers to prepare quality reports and presentations. In addition, the participants get to meet fellow researchers, share experiences and establish network with colleagues from various fields.

Elementary Statistics using SPSS (Statistical Packages for the Social Sciences) for Windows with examples from the sciences is one of the workshops offered in the series. This workshop guides the participants through the fundamentals of using SPSS for Windows as well as to provide basic statistics knowledge necessary for their science-based research needs. The course is structured so as to provide effective training in all the four stages of a typical data analysis process – data definition and input, data modification, data analysis and data presentation. Participants learn to efficiently utilize the power and flexibility of SPSS for Windows so as to get the most from their data. Through a series of "hands-on" sessions and exposure on the basic statistics concepts, they gain that vital experience and knowledge that is able to transform them from SPSS novices to productive data analysts with basic understanding in statistics theory.

The workshop facilitators, Associate Professor Dr. Ling Teck Yee from the Faculty of Resource Science and Technology and Associate Professor Dr. Hong Kian Sam from the Research and Innovation Management Centre, have collaborated to edit the workshop materials into a book. The Research and Innovation Management Centre believes this book can be an important source of reference for academics involved in quantitative research.

The Research and Innovation Management Centre looks forward to more publications arising from the Research Capacity Enhancement workshops which will serve to strengthen the university's efforts to enhance the research and publication activities.

Prof Dr Mustafa Abdul Rahman
Director
Research And Innovation Management Centre

Contents

INTRODUCTION

With the widespread of availability of computers and statistical software and the application of quantitative methods in many fields of studies, introductory statistical course now emphasizes statistical reasoning more than computational skill development. Questions of *how* have given way to questions of *why, when* and *what*? The goal of this book is to enable users to work independently, learning software skills while coming to understand the underlying statistical concepts and techniques.

There are four basic statistical concepts, namely, *population, sample, data collection* and *making inference*. Population is a set of all possible measurement. Sample is a subset of the population. Data collection is the process of collecting data from the sample. Data can be collected in different ways including survey, interview and experimentation. Lastly, we make inferences about the population from the findings obtained based on the data collected from the sample.

Data description and analysis can be greatly assisted by the use of statistical software. This book will cover four stages of a typical elementary data analysis process:

- data definition and input,
- data modification,
- data analysis (descriptive and inferential statistics), and
- data presentation (graph and table).

Inferential statistics are used to draw conclusions and make predictions based on descriptions of data. Inferential statistics involves hypothesis testing and typically involve the following steps:

i. State the null hypothesis, H_o, and the alternative hypothesis, H_a.
ii. Calculate the test statistic value
iii. Look up the critical value or p-value
iv. Make decision
v. Make conclusion

Steps (ii) and (iii) can be achieved using SPSS. Some of the inferential statistics covered in this book are simple linear regression, correlation, one-sample *t* test, independent-sample t test, paired-t test and one-way ANOVA.

CHAPTER ONE

INTRODUCTION TO SPSS

Learning outcomes:

At the end of this chapter, you should be able to:

a) start and end an SPSS session
b) differentiate between data view and variable view
c) name variables
d) understand the use of some of the common tools found in the pull-down menus

---

SPSS stands for Statistical Package for the Social Sciences. SPSS is used for organizing, analyzing and presenting data. All analyses in this book are based on SPSS version 17.0.

1.1 Starting SPSS

To begin SPSS, click **Start** button, select **All Program**, select **SPSS Inc**. folder, select **Statistics 17.0** folder and finally select **SPSS Statistics 17.0 program** as shown in Figure 1.1

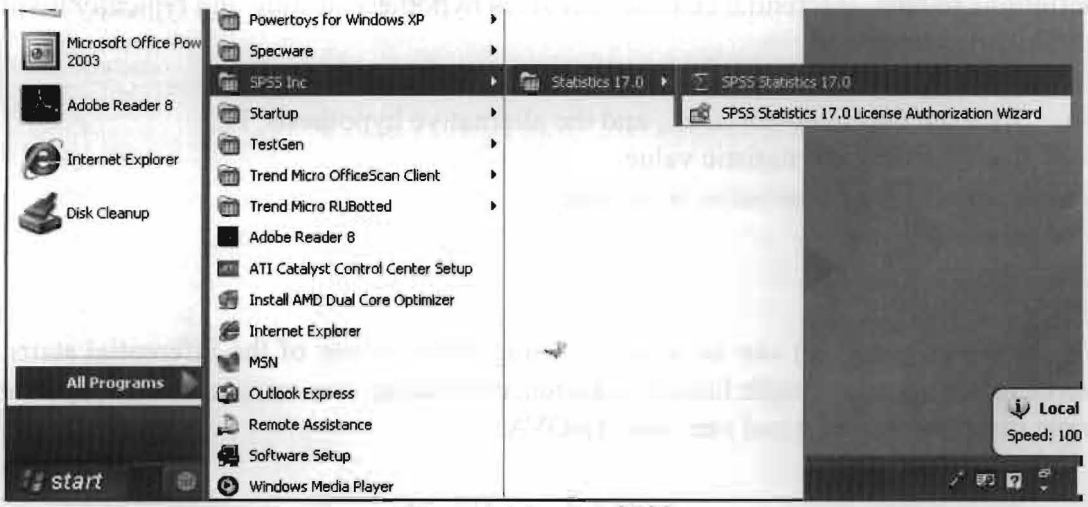


Figure 1.1 Starting SPSS.

Next you will see an **SPSS Statistics 17.0** dialog box. Click the radio button **Type in data**. If you have an existing data file, you can click **Open an existing data source** (Figure 1.2).

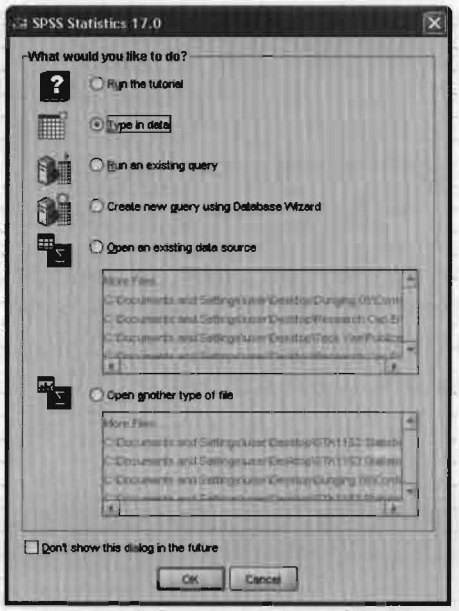


Figure 1.2 Opening screen of SPSS.

Otherwise, click on the radio button **Type in data** and you will see the **Data Editor**. In the Data Editor, you are provided with two views: (a) **Data view** (Figure 1.3) and (b) **Variable View** (Figure 1.4).

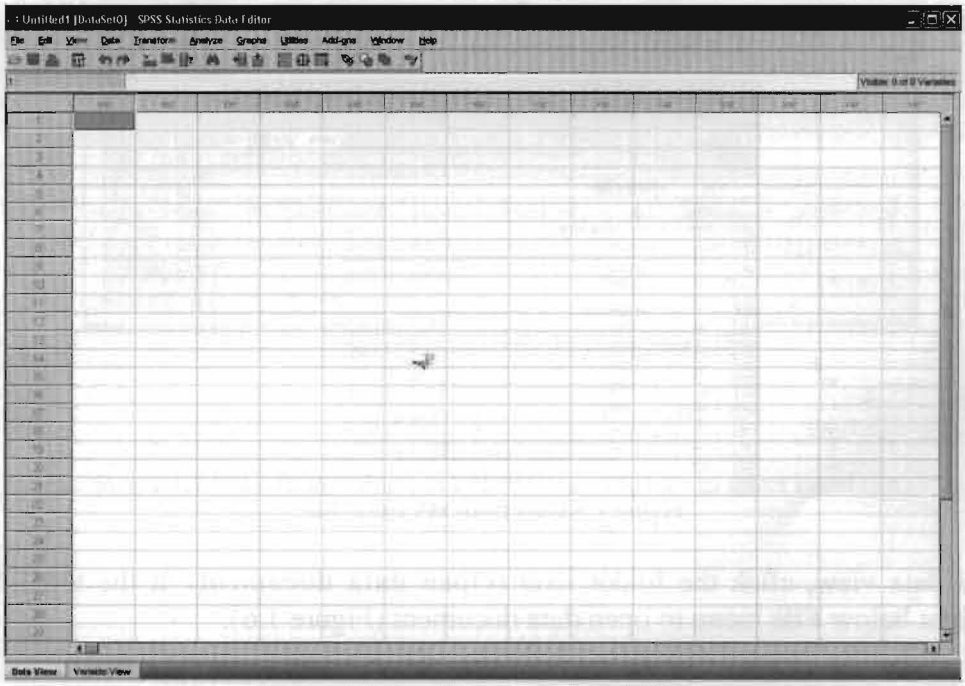


Figure 1.3 Data View in the Data Editor.

Figure 1.4 Variable View in the Data Editor.

The data view is similar to a spreadsheet where you key-in and edit your data. Cases are represented in rows and variables are represented in columns.

1.2 Opening an MS Excel data File from SPSS

If you have data in MS Excel, you can open it in SPSS. For example, you have data in MS Excel as shown in Figure 1.5.

Microsoft Excel - Data for SPSS book

K26

	A	B	C	D	E	F	G	H
1	Day	E. coli Count (CFU/mL)						
2	0	45000000						
3	1	31800000						
4	2	6500000						
5	3	1060000						
6	4	87000						
7	5	48000						
8	6	36000						
9	7	9400						
10	8	3100						
11	9	1820						
12	10	480						
13	11	0						
14								

Figure 1.5 Data from MS Excel file.

In SPSS data view, click the folder icon (**Open data document**) at the top left hand corner right below **File** menu to open data document (Figure 1.6).

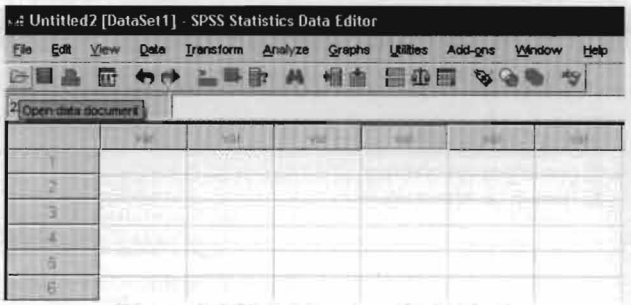


Figure 1.6 To open a data document.

In **Open Data** dialog box, at **Look in:** menu, click the arrow to select your folder where your data are stored. For **Files of Type**, click the arrow and select **Excel (*.xls, *xlsx, *xlsm)** (Figure 1.7). Then, you will be able to see all Excel files in that folder as shown in Figure 1.8. Click the file you want and that file will be shown in the **File name:** box (Figure 1.8).

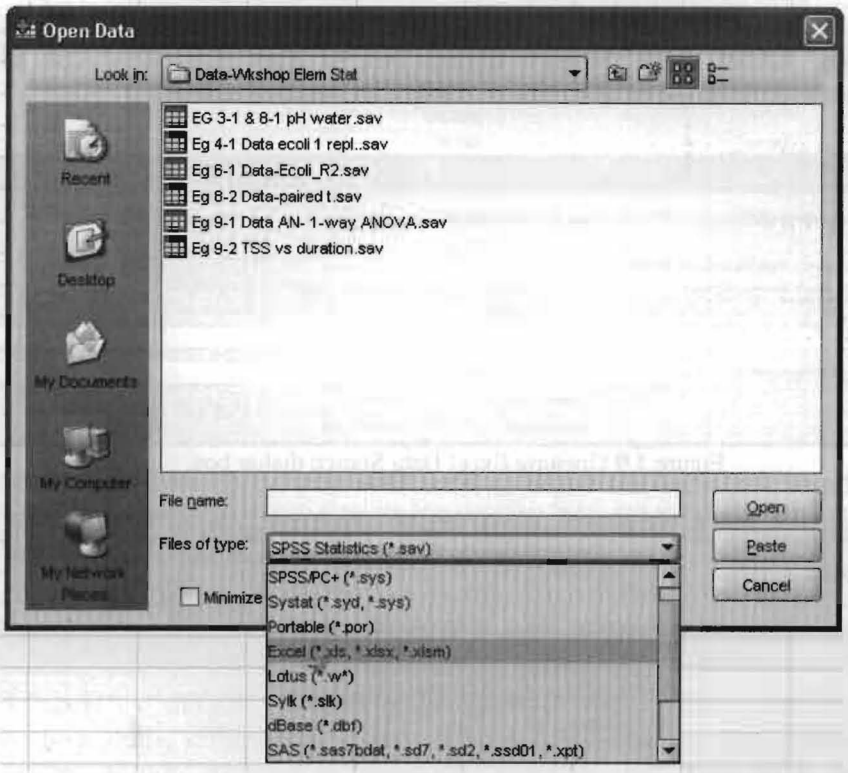


Figure 1.7 Open Data dialog box of SPSS showing how to select Excel file.

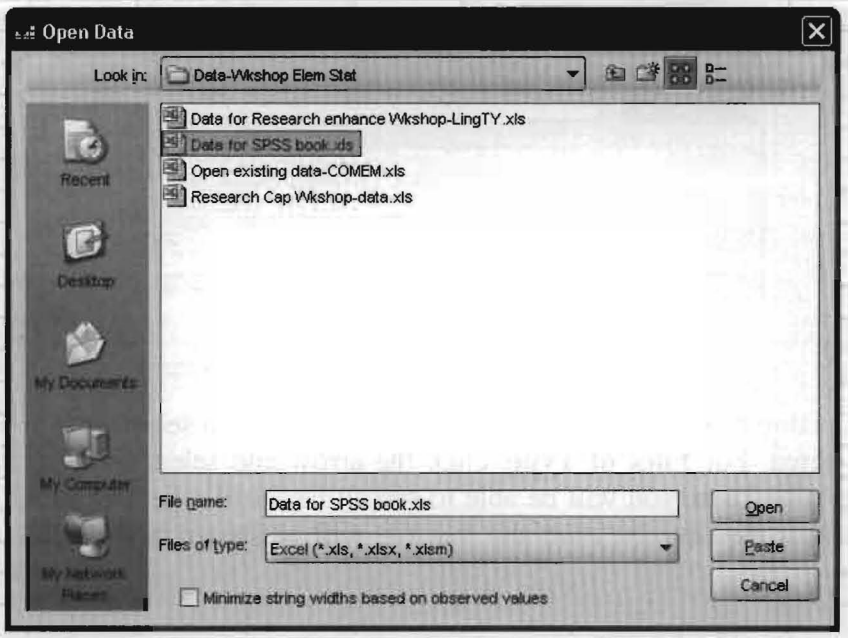


Figure 1.8 Open Data dialog box of SPSS showing Excel file selected.

In the **Opening Excel Data Source** dialog box (Figure 1.9), you can choose the worksheet and the data range of the data you want to analyze.

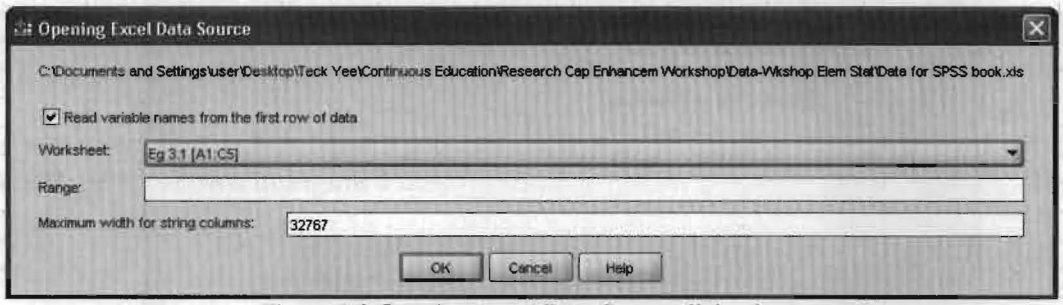


Figure 1.9 Opening Excel Data Source dialog box.

To select the worksheet you want, click the arrow, you will see a list of worksheet in your Excel file (Figure 1.10).

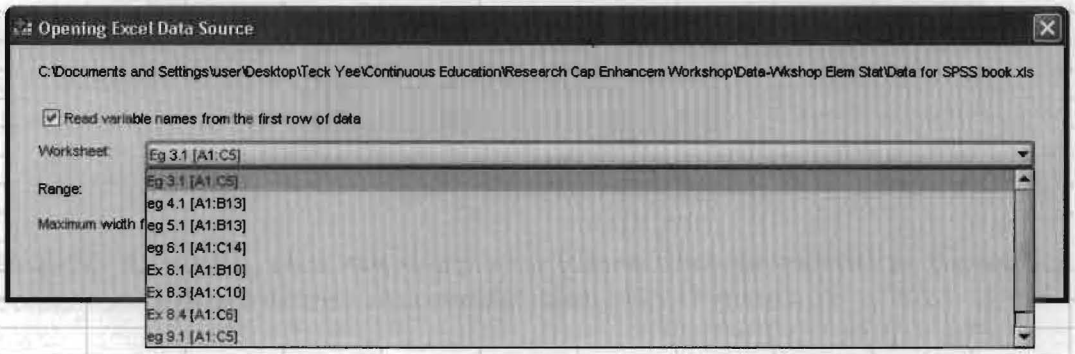


Figure 1.10 Looking for a worksheet in Excel file in Opening Excel Data Source of SPSS.

Selecting "eg 4.1" as the worksheet and type "A1:B10" in range box (Figure 1.11).

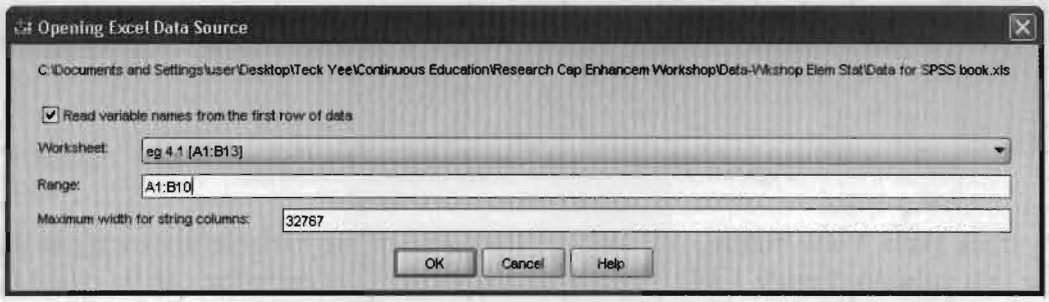
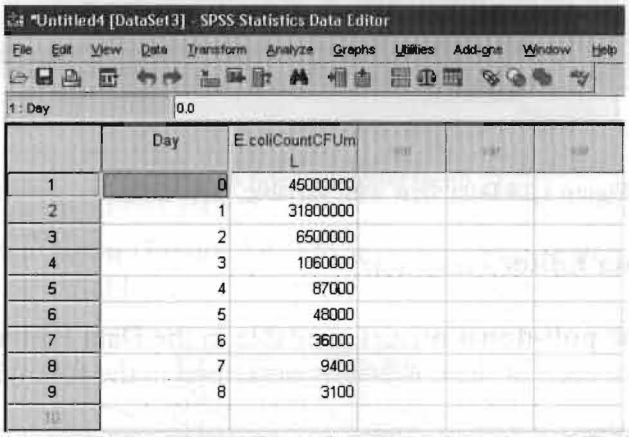


Figure 1.11 Type in the range of data in Excel file to be transferred to SPSS data file.

Click **OK** button, you will see the selected data range in SPSS (Figure 1.12).



	Day	E.coliCountCFUml
1	0	45000000
2	1	31800000
3	2	6500000
4	3	1060000
5	4	87000
6	5	48000
7	6	36000
8	7	9400
9	8	3100

Figure 1.12 Transferred Excel data in the data view of SPSS.

1.3 Name Variables in SPSS

Before proceeding to data entry, you will need to name the variables required for your study. To name the variables, you have to go to the **Variable View** by clicking on the **Variable View** button at the bottom left hand corner.

After you have named the variables, the names will appear at the top row as column heading, replacing **var**. If you do not name the variables and proceed to type in the data from row 1, then by default, they will be given names **VAR0001**, **VAR0002**, ... depending on how many columns of data you have typed in. You can always change the variable name by double-clicking at the name of the variable. To get back to the data view, click the "Data View" button at the bottom left hand corner.

To type in data, you need to name the variables first. For example, you want to name a variable called **Concentration**. Click **Variable View**. In the first cell, you type the word **Concentration** and press **Enter** key on your keyboard. You will see the variable view as shown in Figure 1.13.

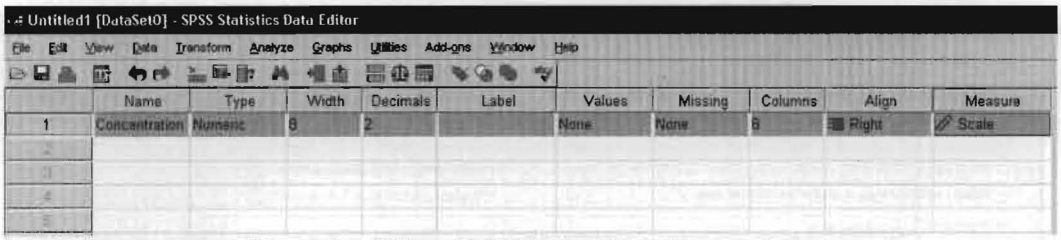


Figure 1.13 The variable, Concentration, is entered.

Now, click **Data View** button, you will see the name of the variable appearing at the top of the first column (Figure 1.14). Then, your values for concentration can be typed in that column. For further details on data preparation, you can refer to Chapter 2.

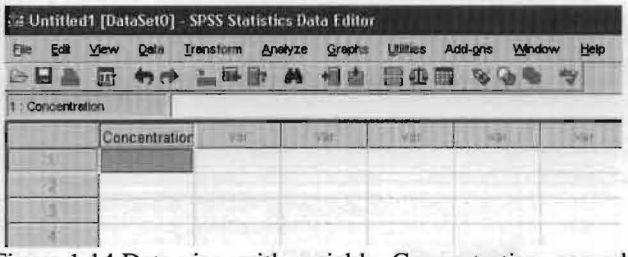


Figure 1.14 Data view with variable, Concentration, named.

1.4 Tools in the Data Editor menu bar

The following are the **pull-down menus** available in the Data Editor and Viewer (Figure 1.15). The function of each of them is briefly described in the following sections.

Figure 1.15 Menu bar for Data Editor

You will use the **File pull-down menu** (Figure 1.16) to create new files, open existing files, save and print files and to exit from the program.

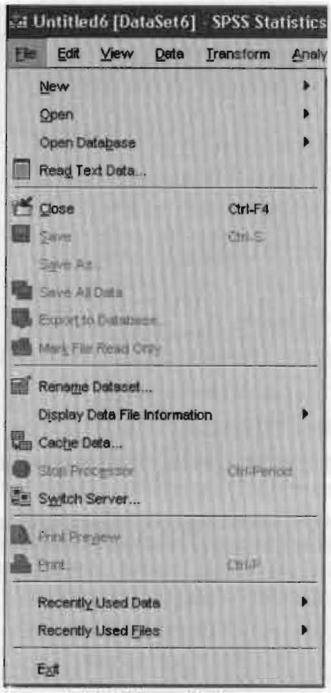


Figure 1.16 File pull-down menu.

Use the **Edit pull-down menu** (Figure 1.17) to modify, copy, search and replace text or data, insert new cases or variables and other preferences options.

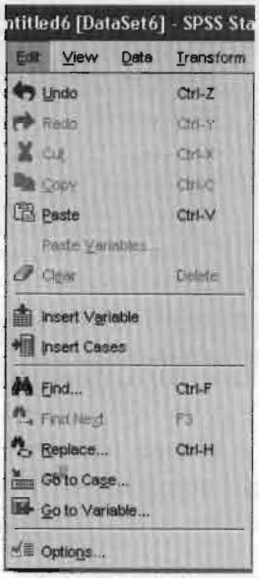


Figure 1.17 Edit pull-down menu.

The **View pull-down menu** (Figure 1.18) is used to activate status bar and toolbars and modify the windows display and select fonts style and size.

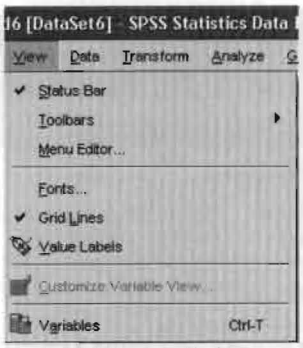


Figure 1.18 View pull-down menu.

You will use the **Data pull-down menu** (Figure 1.19) to define variables and create variable templates. Users can create transpose, sort and select cases and merge files.

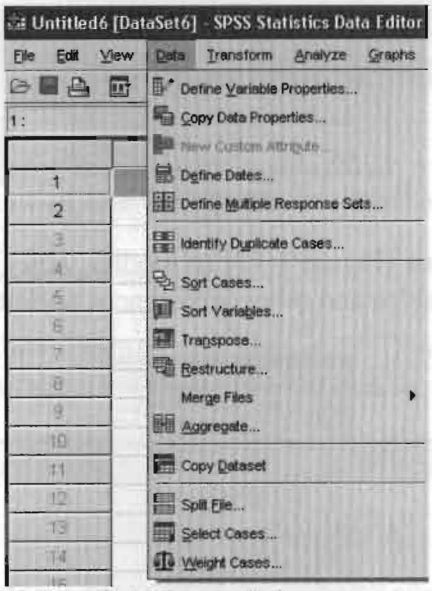


Figure 1.19 Data pull-down menu.

The **Transform pull-down menu** (Figure 1.20) is used to compute new variables from existing variables through mathematical functions and operations, recoding into the same or different variables and rank cases.

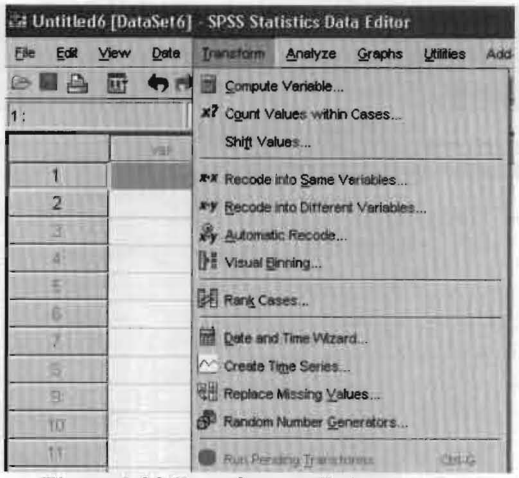


Figure 1.20 Transform pull-down menu.

The **Analyze pull-down menu** (Figure 1.21) provides the means to summarize the data such as frequency, total cases entered, means and analyze the data using different analysis methods according to the design of the experiment such as Paired Samples t-test and ANOVA.

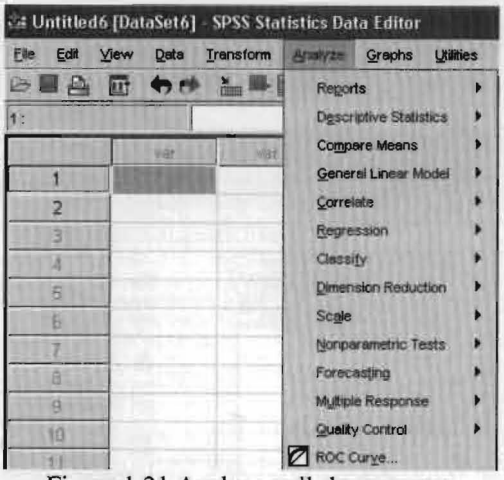


Figure 1.21 Analyze pull-down menu.

If you need to present information from the data in a visual way, use the **Graphs pull-down menu** (Figure 1.22). Different types of graph plots are available for selection from the **Chart Builder** and **Legacy Dialogs**.

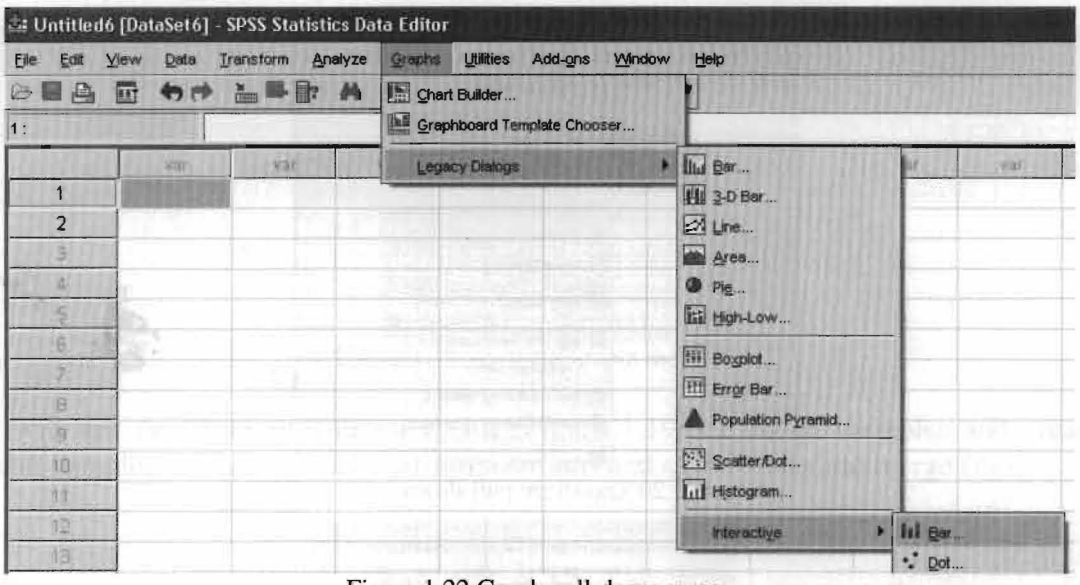


Figure 1.22 Graph pull-down menu.

You will use the **Utilities pull-down menu** (Figure 1.23) to display file and variable information and also define and use different variable sets.

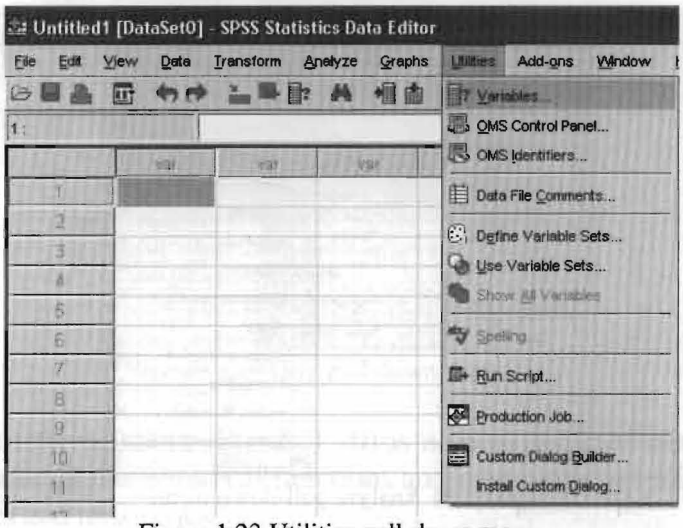


Figure 1.23 Utilities pull-down menu.

The **Window pull-down menu** (Figure 1.24) enables users to move between different files, data and output and control windows such as splitting.

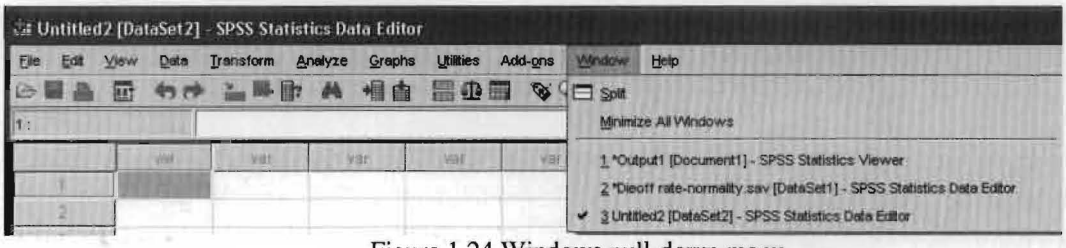


Figure 1.24 Windows pull-down menu.

The **Help pull-down menu** (Figure 1.25) provides tutorials and assistance on how to use the features of SPSS.

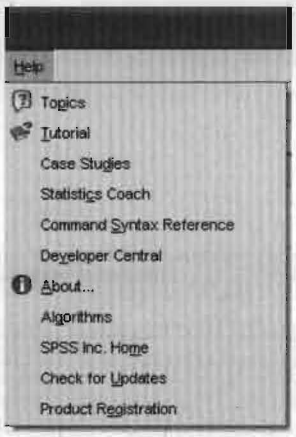


Figure 1.25 Help pull-down menu.

1.5 Viewer in SPSS

The viewer is a separate file for output or results of the analysis. To activate the viewer, click **File**, go to **Open** and select **Output** (Figure 1.26). Then, you will see the **SPSS Statistics Viewer file** (Figure 1.27). It will also open automatically when an analysis is conducted.

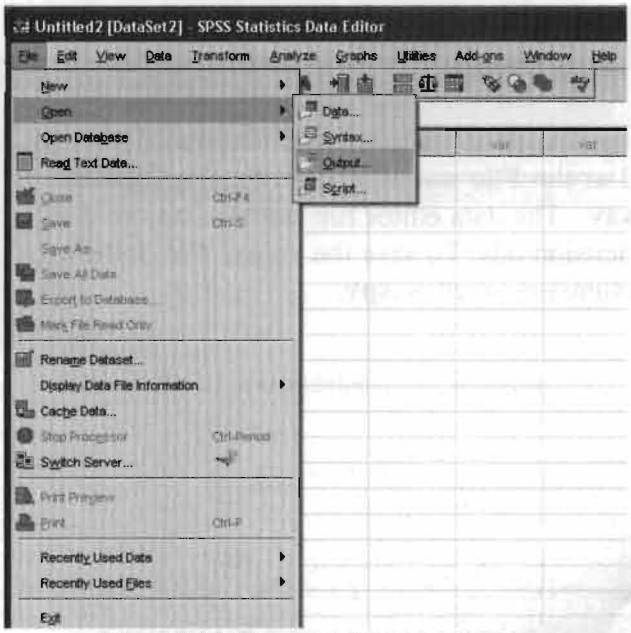


Figure 1.26 How to activate the viewer.

The viewer consists of two panes, the outline pane and the content pane (Figure 1.27). The outline pane contains the outline view of the contents. The content pane contains tables, charts and text of output. You can edit the content by double clicking specific

tables/chart/text. For every analysis, the output will be in this file unless you delete specific output or you close the output file to give way to a new output file.

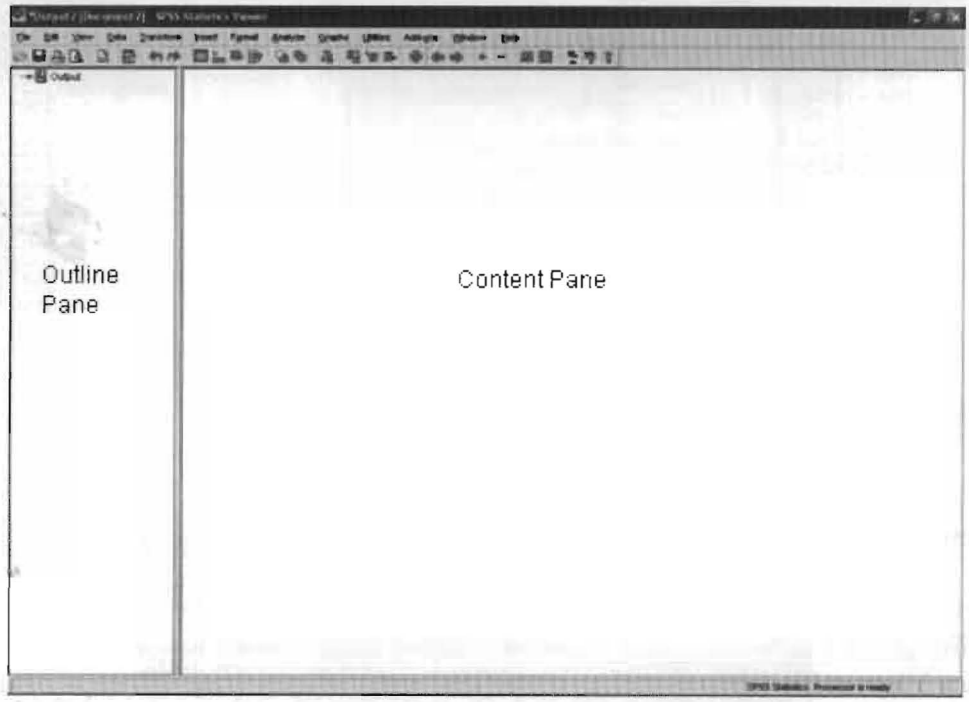


Figure 1.27 Viewer.

1.6 Saving Files in SPSS

To save the data editor file, go to the file menu and select **Save as.** Type the name of the data file to be saved at the **File name** dialog box (Figure 1.28). The file will be saved with file extension **.sav**. The data editor file can also be saved as other types of file such as Excel file with extension **.xls**. To save the viewer file, follow the same step as the data editor file. The file extension saved is **.spv**.

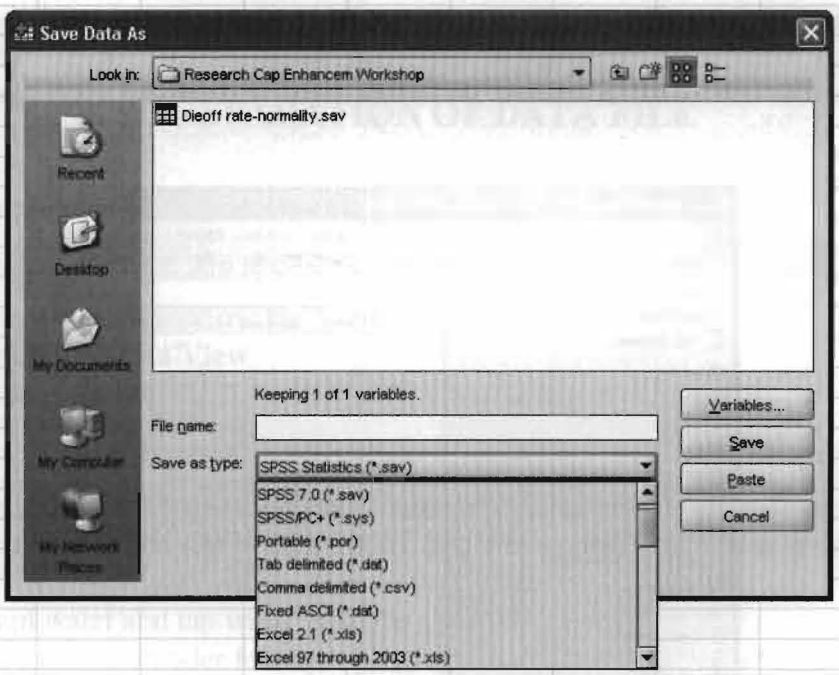


Figure 1.28 Saving data file.

Remember to save your data after data are added or edited and save the viewer file when new analysis is conducted. Another method to save a file is by clicking the **Save this document** icon (second from the left) as shown in Figure 1.29.

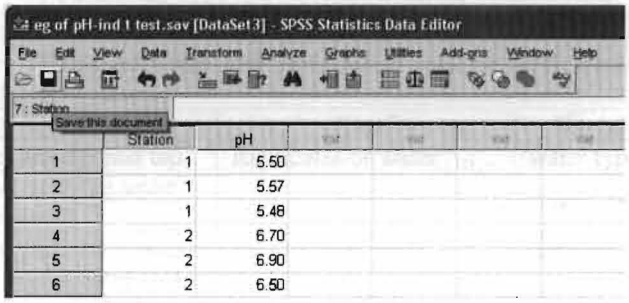


Figure 1.29 To save additional data or output.

1.7 Ending an SPSS Session

When you want to end an SPSS session, go to **File** menu and select **Exit** (Figure 1.30).

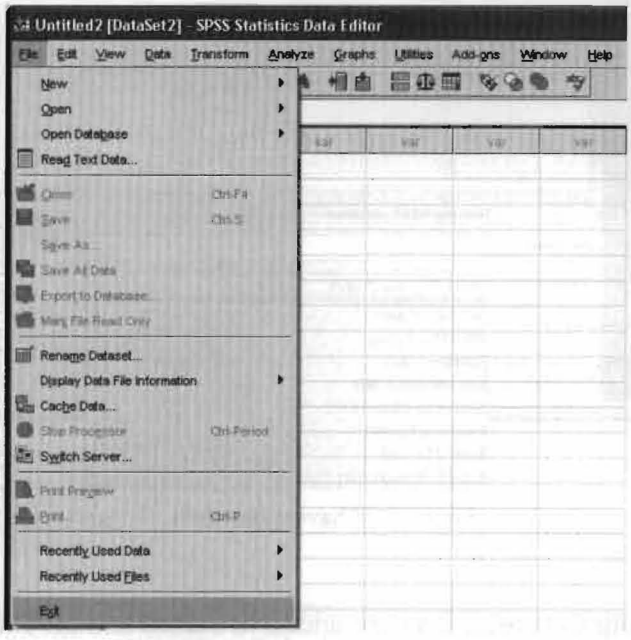


Figure 1.30 To end an SPSS session.

CHAPTER TWO

PREPARATION OF DATA FILE

Learning outcomes:

At the end of this chapter, you should be able to prepare a data file, including to:

a) define variables in **Variable View**
b) enter data in **Data View**
c) save the data file

In a research, the research question states cases of interest and variables upon which these cases are thought to differ (Argyrous, 2005). Some examples of research questions are:

- Do rain water and tap water have the same pH values?
- Is the padi yield under fertilizer A more than that of fertilizer B?
- Is the concentration of phosphorus the same for the upstream and downstream of the discharge point?

After stating the research questions you need to identify the appropriate case and variable. A case is an entity that displays the traits of a variable (Argyrous, 2005). On the other hand, variable is the observation or measurement of the object or subject. For the three research questions stated above, the corresponding "case" and "variable" are shown in Table 2.1.

Table 2.1 Cases and variables of each research question.

No.	Question	Cases	Variables Involved
1	Do rainwater and tap water have the same pH values?	Replicates of water	water type, pH
2	Is the padi yield under fertilizer A more than that of fertilizer B?	Replicates of padi	fertilizer type, yield
3	Is the concentration of phosphorus the same above and below the discharge point?	Replicates of stream water	location, concentration

Data collected can be classified as nominal, ordinal or interval and ratio measurement scale. Nominal data are just categories with no ranking. Examples for animals are birds, cat, fish; for colour: blue, green, yellow, red; for sex: male, female. For ordinal data, they can be ordered or ranked. Examples are short, tall; less than 2, between 2 and 4, greater than 4. Ratio data consists of measurements where the measurement scale has a constant size interval between any adjacent units and there is a zero point with physical meaning on the scale whereas interval data are measurements with constant interval size but no true zero point (Zar, 1996).